Gender Equality in the European Union

A Visualization of the Progress in European Countries: 2013-2020

Exam Paper for Visual Analytics

Students:

Klokholm, Chris	155520
Mealor, Alexander	158801
Ries, Alexander	158292
Torp, Aleksander	158277

Programme:	MSc Business Administration and Data Science
Course Code:	CDSCO1003E
Submission Date:	06.01.23
Nr. of characters:	47174 characters
Nr. of Pages:	25 pages

Table of Contents

Introduction **3**

Data **6**

Data Source 6

Data Cleaning and Processing 7

Challenges and Solutions throughout the cleaning process 9

Index and Index Calculation 10

The Dashboard **12**

Description of the Dashboard 12

Developing and Designing the Dashboard 18

Findings and Discussion **19**

Case Study 1 19

Case Study 2 22

Conclusion **25**

References **27**

Appendix **30**

Appendix 1: Identified datasets, represented key objectives and comments 30

Appendix 2: Outline of the ETL-Process 31

Appendix 3: Example of the applied process guideline to one of the datasets 32

Introduction

Progress within gender equality is an increasingly formalized and measured requirement, on both macro and micro levels in society. Perhaps most ambitious and broad-based of all attempts to formalize this goal is the EU's Gender Equality Strategy, which, in 2020, set out key actions and policy objectives as part of a 5 year plan into 2025 with an aim "to eliminate inequalities, and to promote equality, between men and women" across Europe (European Commission, 2020). The data sets underpinning this dashboard represent country-wide progress in 7 of the 8 key objectives that the European Commission sets out in its Strategy. These are (European Commision, 2020b):

- Ending gender-based violence
- Closing gender gaps in the labor market
- Achieving equal participation across different sectors of the economy
- Addressing the gender pay gap
- Addressing pension gaps
- Closing the gender care gap
- Achieving gender balance in decision-making and in politics

The objective of challenging gender stereotypes has been omitted from the data sets chosen and dashboard produced here, as finding quantifiable and comparable data to represent stereotypes across the EU 27 countries proves extremely difficult.

The Strategy sets out clear actions and targeted measures for the 5 year period, yet measuring progress in these areas is less clearly defined. The dashboard created here intends to serve as one way of illustrating recent trends in progress across the EU 27 countries in each of the target areas, and particularly those that are in greatest need of action. Furthermore, whilst data releases are lagged in many areas, the intention has been to produce a starting point for the 2020-2025 timeframe, which can be readily updated in years to come. Key to assessing progress is a measurement of the 'starting point' in achieving gender equality going into this 5 year period, and both relative and absolute levels of progress across countries in recent years. This remains especially relevant given recent previous strategic engagements such as the 2016-2019 framework, itself defining similar overlapping objectives (European Commision, 2020b). Examining how these previous interventions have impacted gender equality progress can therefore provide insight into 'what works', which areas require an ongoing focus, and the criteria that help to build an effective strategy.

Though one measure of gender equality across the EU already exists in the form of the EIGE's 'Gender Equality Index', its insights and composition offer only one perspective into a complex topic (European Institute for Gender Equality, 2022). But more specifically, the EIGE's insights do not overlap one-for-one with the EU's Gender Equality Strategy as outlined above - the dashboard presented here instead focuses specifically on progress in achieving the EU's key objectives, presenting its own calculated indices and the underlying data and gender gaps in each area. Additionally this dashboard covers a wider perspective, encompassing the 'violence' category which the EIGE measurement does not.

A target audience for the dashboard thus falls under a primary and secondary category. The primary audience consists of the 27 members of the EU Commission, themselves setting the targets, objectives and actions required to achieve progress in equality. It follows that policymakers within EU states themselves - the politicians and heads of state who actually enact the Commission's strategy (and whom the Commission themselves cite as needing to "deepen their engagement" - (European Commission, 2020)) - make up the secondary target audience. Both have an interest in monitoring progress, but their needs are different. Commission members have a need to formulate an appropriate overall strategy with effective actions and actionable targets, whilst EU state policymakers have a need to ensure their own implementation of the Strategy is working with actionable insights. Commission members are likely more interested in how close the EU as a whole is towards achieving gender equality. State policymakers likely care more about their own country's performance relative to others. The dashboard itself caters for both of these needs by illustrating broad EU progress in terms of an overall gender equality index and sub-indices across the objectives, and country-specific progress in terms of the trends in the data underlying these indices. Both audiences have overlapping interests in monitoring trends in progress over time and relative to other key time periods (e.g. 2016 - 2019). Their reasons for wanting such insights too differ. As the Commission itself describes - "The promotion of equality between women and men is a task for the Union, in all its activities, required by the Treaties. Gender equality is a core value of the EU, a fundamental right and key principle of the European Pillar of Social Rights" (European Commission, 2020). For individual states, societal and economic improvements in areas such as "jobs and higher productivity" is possibly even more key.

Both trends and snapshots in progress across the key objectives matter for the primary and secondary target audiences. Trends allow both groups to examine whether shifts in equality are pervasive and lasting or temporary - one relevant element of this is the analysis of policy interventions, such as the introduction of a hypothetical minimum wage agreement, which can be more readily analyzed to check for longer-term effectiveness in improving economic equality across genders. In our *Findings* section we run through this specific example, looking at the introduction of the German Minimum Wage Act in 2015 as a case study and potential driver of progress. For Commision members in particular it also allows for the analysis of previous strategy efforts, key for informing future strategy formulation. Snapshots - whilst important in analyzing the latest measure of progress and achievement - arguably only form part of the overall picture, since the nature of progress in gender equality is a slower-moving one requiring a broader viewpoint. For this dashboard, the 2020 snapshot however provides a key benchmark going into the 5 year period, and upcoming years will still help to inform decision-making.

The exact form of the decision-making or implications brought about by insights of the dashboard clearly differ by audience. For EU Commission members, the expectation is that monitoring trends in progress across each objective, before and after the strategy intervention, will go towards informing future strategy efforts. In this sense, a potential 2025-2030 strategy - its objectives, targets and actions - will be led by the insights from the current period. In the negative sense, targets that have proven unrealistic may be reframed, actions that have been ineffective can be tweaked, and objectives that do not capture the reality of gender inequality expressed in the data can be dropped. In the positive sense, targets that have been surpassed can be revised to more ambitious levels, effective actions and policy measures can be applied elsewhere or more broadly, and objectives that have captured key nuances of gender inequality can be focused on in future strategies. Examining progress on a country by country basis can also allow the Commision to take more granular actions on specific states, perhaps issuing warnings or specific recommendations to those lagging or excelling in certain areas of gender equality. For EU state policymakers, the insights inform specific implementation within their own countries - the ability for states to compare their own progress against others (perhaps neighbors or regional counterparts) opens up space for policy collaboration or, at minimum, policy-making inspiration from what has worked for others.

It is important to emphasize that the insights generated from this dashboard should make up only part of a broader analysis of gender equality across the EU. Given the overall complexity of the topic, a general lack of alternatives specifically designed for measurement of gender equality progress vs the EU's own Strategy targets, and the need for both broad-based EU and country-specific granularity, the insights generated from this dashboard should be valuable to policymakers at both of our target levels.

The dashboard can be accessed here:
https://public.tableau.com/app/profile/aleksander.torp/viz/EUGenderEquality2013-20201_0/Dashboard1-USETHIS

Data

Using correct, reliable and clean data is essential to provide the target audience a high-quality dashboard. For this, data sources have to be determined and the necessary data sets specified. Next, the data has to be cleaned and processed in a way so that the data forms a solid basis to visualize it. In the following sections the complete process including finding a data source, the data cleaning and necessary calculations are described. Furthermore, an outline of the ETL process can be found in Appendix 2.

Data Source

There are numerous databases or datasets freely accessible on the internet providing gender statistics that could have been used for displaying the situation of the aforementioned key objectives from EU's Gender Equality Strategy. To ensure that data sources of high quality and reliability are used, a first focus was on online databases from reputable institutions such as the World Bank, OECD, the statistical office of the European Union (Eurostat) or the European Institute for Gender Equality (EIGE). After a more detailed exploration of these databases, Eurostat was determined to be the main data source, giving us the datasets that underlie the dashboard, for the following reasons:
The databases from World Bank and OECD offer a variety of datasets for most of the key objectives such as information about the gender wage gap or gender employment gap (The World Bank Group, 2023 and Organisation for Economic Co-operation and Development, 2022). Nevertheless, the necessary data for some of the EU27 countries is not all available for a period over the last ten years in the World Bank's database (The World Bank Group, 2023b) and the OECD already has missing data in data sets that should be easier to collect in terms of availability and transparency, such as the gender wage gap (Organisation for Economic Co-operation and Development, 2022). The choice of Eurostat instead of EIGE as the main data source comes from the fact that EIGE takes Eurostat as a source for many of its statistics (e.g. European Institute for Gender Equality, 2017). In addition, EIGE data is often already manipulated and presented for a specific purpose. Eurostat is therefore a more 'neutral' data source. For the dashboard, it is essential that the data basis is founded upon neutrally collected data without a certain statement intention. Despite this, EIGE also serves as a data source for one of the datasets. This is because the data on national parliament members were collected by Eurostat only up to 2017, and EIGE took over this data collection since then (European Institute for Gender Equality, 2022b). Both Eurostat and EIGE provide metadata for their datasets containing detailed information in general

and about the dataset, such as the quality, the relevance and the accuracy of the data as well as the reliability of the data in the dataset (Eurostat, 2021).

The data for representation and the ways of assessing the key objectives from the EU's Gender Equality Strategy differ significantly. There is no single dataset that includes the ideal data for every objective at once. Hence, the goal was to find the most suitable data set for each key objective from the EU's Gender Equality Strategy in order to represent the status and development of these objectives in a Key Performance Indicator (KPI). From the data sources Eurostat and EIGE, we identified a total of seven datasets that form the data basis of the dashboard. Officially the Gender Equality Strategy counts 8 key objectives. We were not able to find a suitable dataset that provided adequate information to represent the objective “Challenging gender stereotypes”. Therefore, this objective is not represented in the dashboard or included in any form in the data cleaning and manipulation process. According to the metadata, some of the datasets may contain data with low reliability or have missing values for one specific country, year or another feature in the dataset. The following cleaning and data manipulation process is aimed at dealing with these problems. In the case that this process is not able to avoid an issue with the dataset (i.e. missing data), the dashboard will clearly show and point out these issues so that the user is aware of that and can interpret the data correctly. The datasets, their sources, which key objective they represent and further comments regarding the expected challenges throughout the cleaning process are listed in a table in Appendix 1.

Data Cleaning and Processing

To extract and shape the data for the purpose of creating a proper dashboard for the target audience, cleaning, manipulating and processing the data is essential. To perform those steps, specific tools were used and a cleaning process guideline was introduced. The general way of cleaning the dataset included exploratory data analysis, data cleaning and manipulation, and specific data calculations. In the end, the cleaned data sets should be structured in a way so that each variable forms a column, each observation forms a row and each type of observational unit forms a table (Wickham, 2014). In this case, the observational units are the seven key objectives of the EU's Gender Equality Strategy.

Tools

The programming language Python was used to perform the data cleaning process and other calculations in order to create a solid data base for the dashboard. For collaboration, the version-control platform GitHub was used. This platform centralized the cleaning process in one place to make sure that everyone was working on the most recent code and with the up-to-date datasets.

Cleaning Process Guideline

Since more than one data set needs to be cleaned, a process guideline was defined first. This guideline is an orientation for the person who wants to clean a dataset and establish a certain cleaning standard that each dataset must go through at least. The process guideline included the following steps:

1. Exploratory Data Analysis and initial cleaning steps

In this first step a short Exploratory Data Analysis (EDA) is performed. EDA means looking at the data to understand what might be interesting to visualize, but also what has to be done during the cleaning process with the data set (Knaflic, 2015). We can explore the features of the dataset, the values in the dataset and get a first impression of what to do in the cleaning process. Besides that, unnecessary columns for the further process have been identified and deleted. To ensure that every dataset has named the countries in the same way, we created another dataset containing the country names that we wanted to use in the dashboard and mapped them to the dataset in the cleaning process. If necessary, the data sets had been reshaped according to Wickham so that the columns have variable names as headers and values only in the rows, only one variable is stored in each column and the variables are only stored in columns (2014).

2. Null values and invalid rows

In the second step, the null values or invalid rows should be handled. Since the datasets can differ considerably, there is not one uniform correct way to handle those values. The different approaches are described as follows:

The datasets on gender representation across economic sectors and the gender pay gap consists of multiple sectors per country for any given year. To handle missing values for one or more sectors, the values were added up and divided by the number of sectors with data for every year, creating an average to represent the given country's performance, regardless of whether some sectors lacked data. After performing the calculations, missing values were replaced with zeros to maintain a consistent number of rows and columns. Concerning the pension gap dataset, there was only one missing value, and since no calculation was necessary, it was replaced with zero after creating the index. Null values in the violence data sets were replaced with zeros since this allowed for easier filtering later when visualizing, and suitable replacements (means, previous numbers) were not a fair reflection when ultimately calculating indices. When calculating the average violence figures between the three violence types, the number of non-null values was taken as the denominator. Care data was similarly handled by replacing null values with zeros for the same reason. Since the data was split into male and female figures, and a gap was calculated between the two, the logic that if either value was zero then the overall figure would be left at zero was applied. This prevented misleading gaps from being generated. Before going over to the next step, a short quality check has been established here. It is reviewed if all countries in the dataset have the same number of rows in the dataset, so that there is no missing data for each country. Since every dataset contains time data, another check was if all years have the same number of rows as well ensuring that a year does not have missing data.

3. Further individual cleaning

Further individual cleaning is the third step. However, the individual cleaning can happen anytime in the cleaning process because it could be necessary before another specific step. For example, it makes sense to transform the time series data to a yearly basis before performing the quality check in step 2 of the cleaning process. The most relevant and challenging individual cleaning actions are described in the next section “Decisions and Challenges throughout the cleaning process”.

4. Index calculation

The final step for the data cleaning and processing - each objective or data set needs to have its own index. Therefore, an index value is calculated for each row in the datasets providing the basis for the following overall index calculation.

A code snippet of the applied process guideline to one of the data sets is provided in the Appendix 3. To review complete code, the raw datasets and the cleaned datasets please refer to the GitHub repository https://github.com/Reese181/VA_Dashboard.git.

Challenges and Solutions throughout the cleaning process

The structure of the datasets was little to very different. The cleaning process has been adapted to each one of them. Nevertheless, some challenges arose in the cleaning process and decisions have been made to solve them.

Null Values

Some of the data sets had far more null values than others. The null values were replaced with zeros in all datasets (except one). Zeros can be shown as break or blank in the data visualization tool Tableau so that they are clearly recognizable by the user. Furthermore the zeros do not impact the index calculations, by applying a workaround to ignore them. One different approach of handling null values has been applied in the dataset “gender representation across economic sectors”, as described in the section “Cleaning process guideline”.

Calculations

Calculations have been made on each dataset individually. The most common calculations were *gap subtractions*. The most datasets contained values for females and males. Since some of the objectives addressed gender gaps, we needed to subtract the male values from the female values to generate the gap values. In other cases we needed to calculate averages since the data was distributed on different features in the dataset, for example across different economic sectors (Eurostat, 2022c).

Timeframe

The timeframe of the data varied across the datasets. Furthermore, the quality of data was poor in some years (too many null values, complete blank rows, missing countries) so that we needed to find a timeframe in which every dataset could display a proper amount of data points and in the right quality. This was necessary for both calculating the overall index and for the data visualization.

Index and Index Calculation

A key element of the dashboard is the inclusion of an ‘Overall Index’ of gender equality, derived from the measures that assess progress towards 7 of the 8 key goals set out by the EU’s Gender Equality Strategy. Inspired largely by the Human Development Index in both its use and calculation, in the same vein, the index measure is taken as a “summary composite measure of a country’s average achievements” (World Health Organization, 2022). For our use case in particular, an index and its sub-indices help to quantify relative performance both over time and between countries in a more objective manner than individual data points alone may produce. This is especially relevant for our datasets given much of the information is captured in gap form or more complex population based measures such as incidents per 100k inhabitants. Leaving our target audiences to garner insights from the underlying data alone would therefore not suffice if decision making could not reflect a broader view with the option to then dig in to the underlying information.

The Overall Index figure is formed from 7 underlying sub-indices - Violence, Care, EcoSector, Employment, Pay, DecisionMakers and Pension. Each of these sub-indices is calculated in a similar way overall, but with unique values tailored to each dataset. Drawing on the calculation process of the Human Development Index, the 'Dimension index' formula is applied to the underlying datasets (Human Development Report, 2016, p. 2).

$$\text{Dimension index} = \frac{\text{actual value - minimum value}}{\text{maximum value - minimum value}}$$

Minimum values and maximum values consist of 'goalposts' which allow underlying data to be normalized to a relative 0 to 1 scale, where 0 represents perfect inequality, and 1 reflects perfect equality. These values differ based upon the data set in question. For example, in any categories that rely on a measured percentage gap between females and males, 0 - corresponding to no gap - is naturally the 'best' or 'maximum' value, whilst 100 is the 'worst' or 'minimum' value. For data that relies on absolute numbers, such as the violence category, the maximum value again corresponds to 0 (in this case, 0 female victims of violence), whilst the minimum value corresponds to an average of the maximum values of each of the underlying data sets (homicide, sexual assault and rape). Each of these sub-index figures is displayed on the front page of the dashboard.

The calculation process continues to follow that of the HDI in the actual aggregation of the sub-indices to derive the overall index. This involves taking the geometric mean of the sub-indices (Human Development Report, 2016, p. 2).

$$\text{OverallIndex} = (I_{\text{Violence}} \times\ I_{\text{Care}} \times\ I_{\text{EcoSector}} \times\ I_{\text{Employment}} \times\ I_{\text{Pay}} \times\ I_{\text{DecisionMakers}} \times\ I_{\text{Pension}}\,)^{\frac{1}{7}}$$

The geometric mean is used here as a method of managing time series data that at points exhibits serial correlation, whilst also giving "higher importance to the dimension having lower performance, and penalizes unbalanced development" (Mishra & Nathan, 2013, p. 5). This arguably more conservative approach also fits the spirit of achieving progress in gender equality - that all elements have to improve to make a meaningful difference to society. Geometric means are taken with respect to all non-null data points to ensure the measure does not punish countries with less data, so the product is raised to the power of 1/non-null values.

Formation of the index also impacted the time-frame that was settled on in producing the dashboard, since finding overlapping dates for all 7 data sets in terms of availability and minimizing the prominence of null values meant that 2013-2020 was the chosen range for calculation and wider display.

As previously mentioned, it should be noted that the EIGE's Gender Equality Index does already exist, yet its calculation and underlying data sources differ from the index produced here. Notably, EIGE's version does not include the violence category which is catered for here - it cites "a lack of evidence to assess violence against women" (European Institute for Gender Equality, 2022c). Yet even with some shortcomings in comparability and consistency of the underlying data, our method of specifically looking at female victims, and averaging out numbers across crime categories provides a different way of examining these statistics. The Overall Index and its sub-indices also rely on different data sources and consider different categories of gender equality. Essentially, the Overall Index is a specific measure of progress in achieving the EU's Gender Equality Strategy goals, while the EIGE's index is a broader measure of gender equality.

The Dashboard

For the construction and publishing of the dashboard, we used the software Tableau, primarily due to the software's functionalities, its accompanying publishing features, and for the possibility of creating a map-based dashboard. The dashboard is made available using Tableau Public, with a small window size to increase the accessibility to anyone with a computer and internet access. The following sections explain its features and elaborates on the reasoning behind the design choices

Description of the Dashboard

The dashboard consists of two views, one for displaying aggregate performance EU wide and one for showing information on selected countries of interest.

Aggregate View

The main page displays the larger overview of how the EU as a whole is performing on all gender equality indices, according to the overall index scores, as well as how individual states are doing relative to each other (Figure 1).

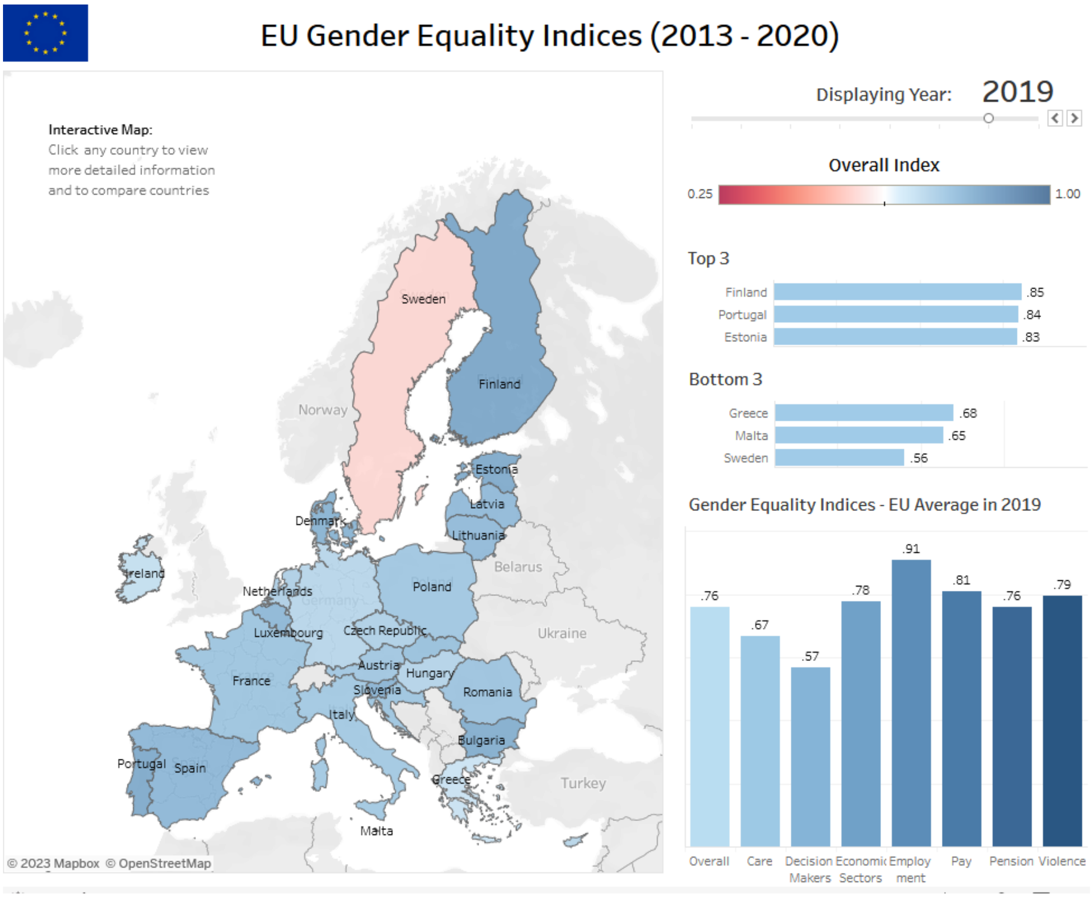


Figure 1: Aggregate view of the dashboard

The relative performance of every country is presented as a heatmap on the left side. A color palette ranging from red (worse status) through white to blue (better status) presents the countries against a neutral background. This is used to create a contrast that guides the eyes of the reader and to create emphasis on the relative magnitudes of the scores..

In the right upper corner (Figure 2), a slider enables the user to change the year, so as to examine the aggregate development across the EU over time. Below the slider, one finds the gradient explaining the intensity of the coloring and the overall corresponding index score, with a higher overall index score presenting smaller gaps between men and women and indicating better gender equality. Next, one sees the top and bottom performers in terms of the overall score. This is presented to both show the distance between best and worst, and to draw the user's attention to countries of particular interest in terms of the selective view to obtain a greater understanding of objectives to investigate.

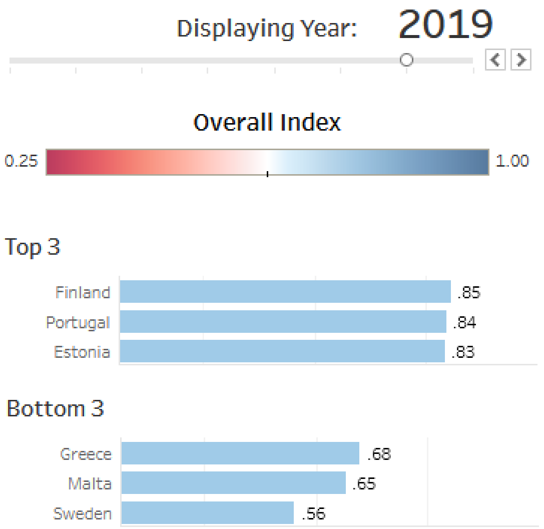
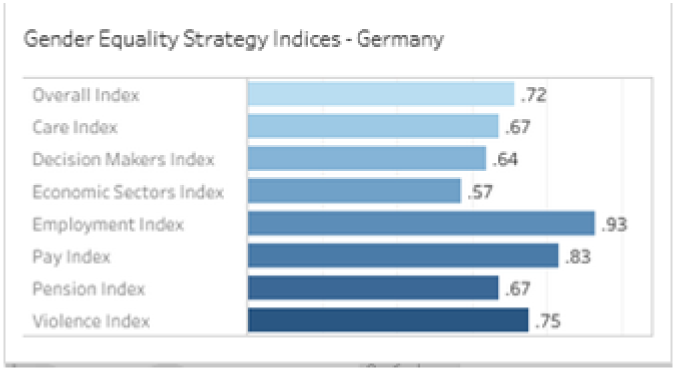


Figure 2: Upper right-hand corner of aggregate view - year slider, Overall Index color gradient and top/bottom performers

Next, by hovering over a country on the map, its individual scores in the overall index and its sub-indices are displayed. A bar chart is used to convey relative distances on its performance, where higher scores equal smaller gender gaps, (Figure 3), and, together with a similar coloring gradient, one can easily compare country stats with EU wide averages found in the bottom right corner (Figure 4).



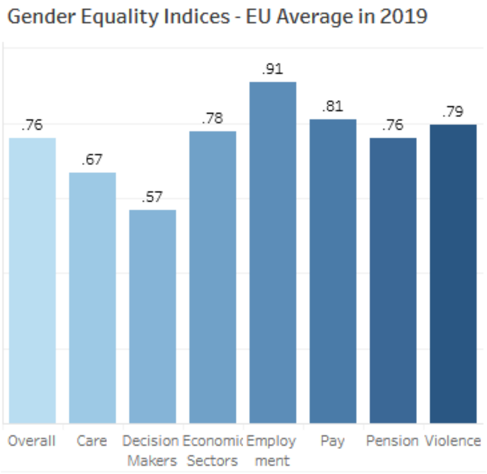


Figure 3: Sub-indices by country level (Germany)

Figure 4: Sub-indices by EU aggregate level

Textual information is presented with a color palette of black and grey throughout the dashboard. Headlines are bolded and colored black and other text and numbers are in grey, in order to lead the readers eyes to the information in a natural order and to avoid visual noise. The bars are labeled with the indices scores to avoid using vertical axes to conserve space. Furthermore, with all the scores ranging between 0.00 and 1.00, for increased readability and quicker interpretation, the leading zero is not displayed.

The Selective View

The selective view is reached when the user clicks on any country on the map. One is then presented with the selected country's scores on the seven equality objectives. The initial view intentionally states the name of the country for which data is presented in a larger font size to avoid misunderstandings (Figure 5). Through using a combination of line and bar charts, users are informed on both the country's situation in each individual metric by absolute figures as well as on the direction of developments in gender differences.

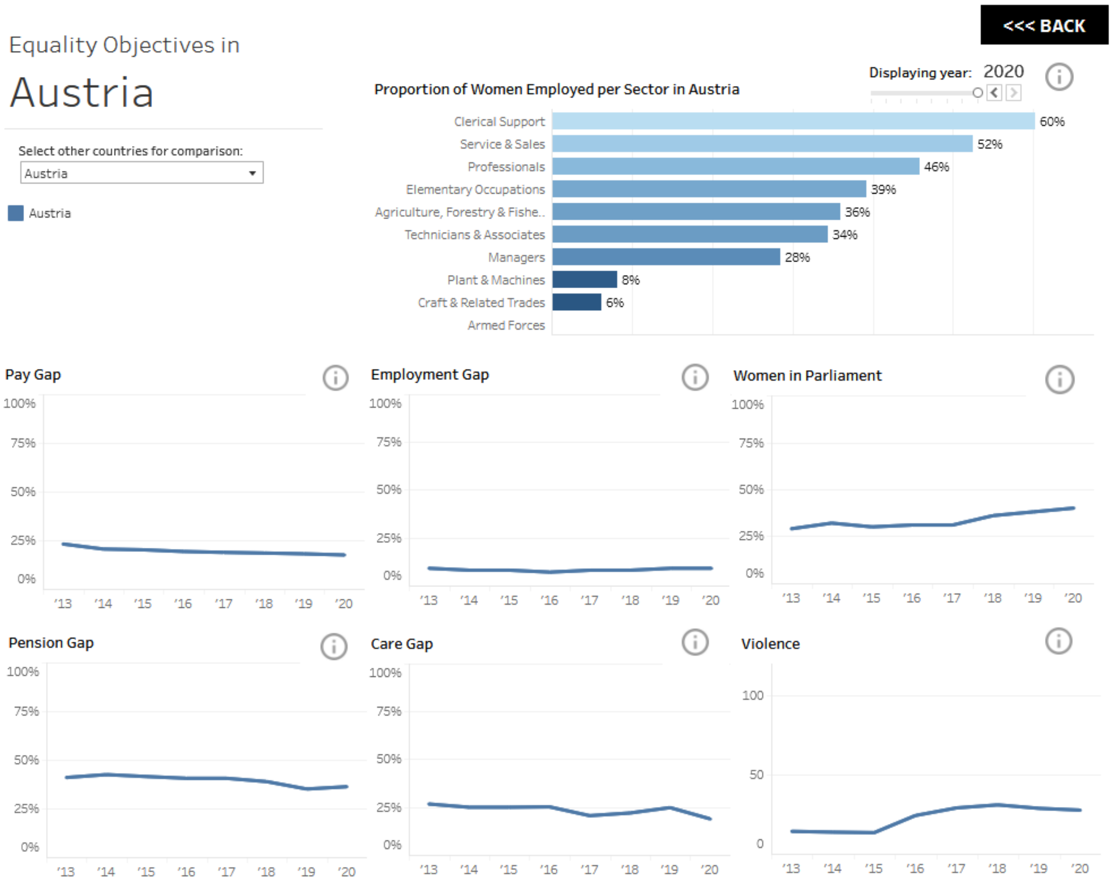


Figure 5: Selective view in dashboard - Austria

The axes of the timeline charts use fixed ranges, 0-100% and intervals of 25%, to allow for cleaner and unison comparisons both between the gaps and to visually convey the differences in the various objectives, both between objectives and countries (Figure 6). The violence measure - where the range is set to the maximum average value across all countries for the vertical axis - is the one exception to this. When there are missing data in the time series, this is shown as breaks in the lines to clearly inform the user that there is not sufficient data available. By hovering over any of the charts, one sees the absolute values for a given year. The EU's objective on equal compensation encompasses pay and pension, but we split this into two to provide greater detail. Additionally, for every chart, an informational button is placed in the upper right corner to provide the user with more details on what statistics the chart represents and the underlying data.

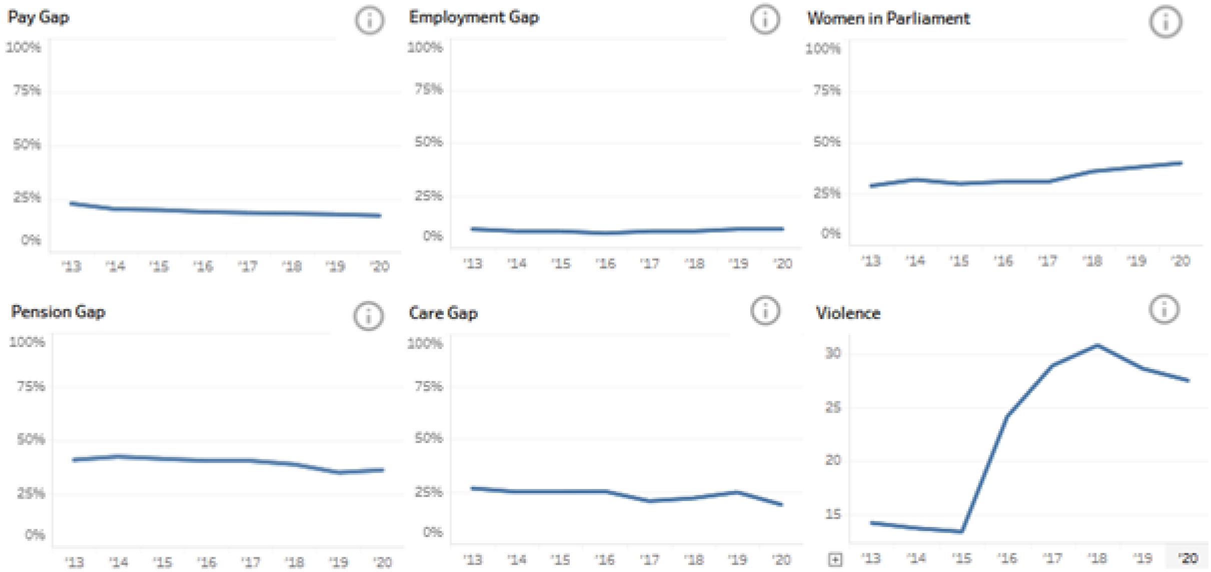


Figure 6: Timeline charts for equality objectives

The gender distribution across economic sectors is presented using bars to show the proportions between sectors, (Figure 7), and also due to the nature of the data. A slider enables users to examine development of gender distributions over time with the sectors sorted in descending order by highest female percentage. Darker coloring is used to draw the user's attention to the sectors with the lowest female participation.

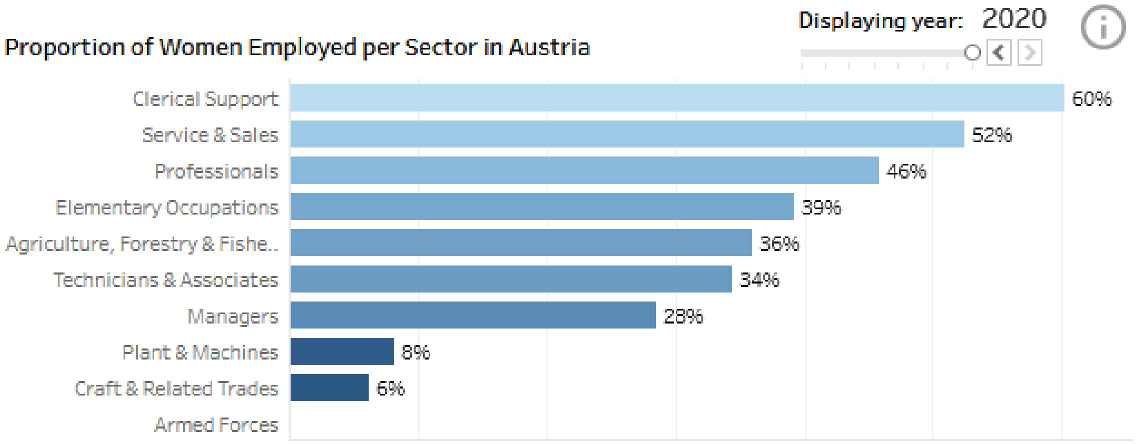


Figure 7: Economic sectors bar chart (Austria)

The selective view is designed to accommodate the inspection of both individual and multiple countries. Using the dropdown menu in the upper left corner, one can select unlimited countries for comparison. Each country's statistics are then displayed in the same charts and distinguished from each other by colors. Due to the large number of countries available for comparison, a large color palette is necessary to use to properly differentiate and the tooltip informs about which country a line represents, as shown in Figure 8.

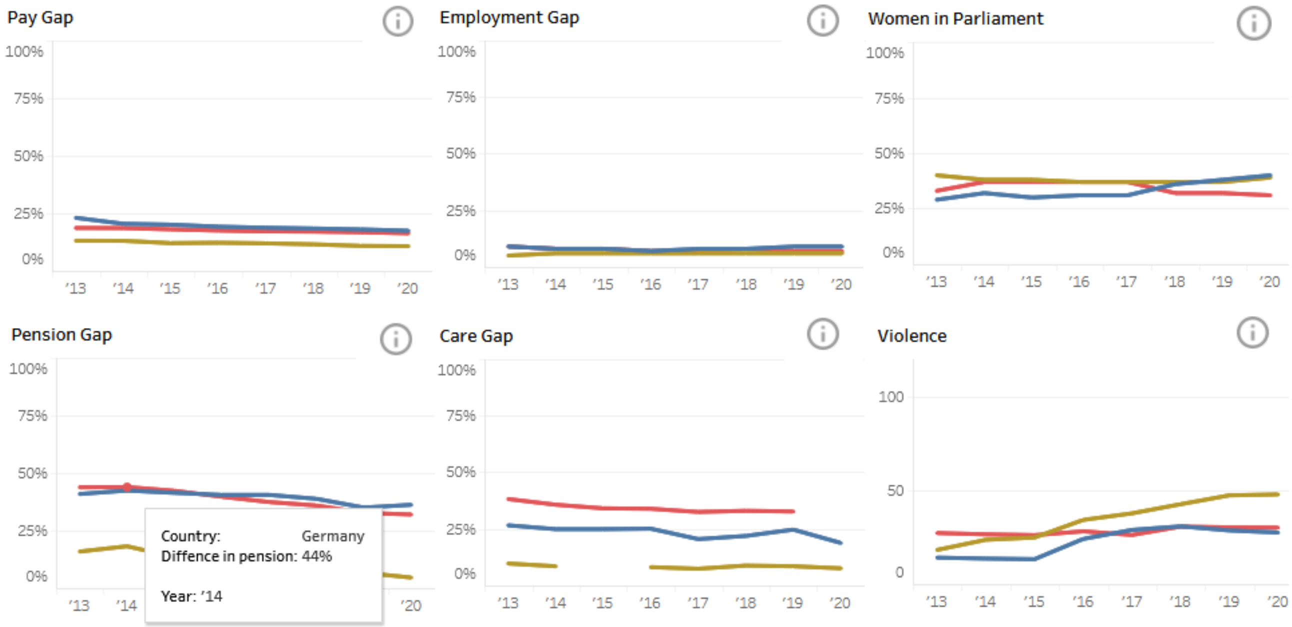


Figure 8: Country comparison via selective view

Developing and Designing the Dashboard

The initial idea for the dashboard remained the same throughout the whole process, from draft to final product, of presenting the status of gender equality objectives from two perspectives, one aggregate view and one more in depth for countries of interest. This came about in the preliminary review of the datasources, by first reflecting on who are the audiences of the dashboard, what kinds of information the receivers need and how is that best conveyed, as outlined by Knaflic (2015). Given the target audience, ranging from decision makers at national levels up to EU, it made the option to toggle between the two views reasonable.

It has been an iterative process of testing different visualizations of the data and reducing the visual elements to the bare minimum to achieve a dashboard that "...speaks clearly and immediately" to the audience (Few, 2006, p. 12). The first draft consisted of working out primarily the technical aspects in Tableau and testing different layouts and visualizations of the data. Both placement and alignment objects are rarely noticed when done properly, but a major distraction when elements are out of place (Knaflic, 2015). Hence, the second round focused on fitting the dashboard elements together in a coherent manner, with respect to the zig-zagging reading movement of people from left to right. Furthermore, "Display mechanisms that clearly state their message without taking up much space are required" (Few, 2006, p. 27) to take advantage of the limited space available. Therefore, the selective view, for example, was developed with the aim of presenting multiple layers of absolute values to allow for comparison both across the objectives and between countries by using fixed ranges across chart axes. Considering that a dashboards purpose is to enable the reader "to quickly point out that something deserves your attention and might require action" (Few, 2006, p. 27), preattentive attributes are used to guide the attention of the user (Knaflic, 2015). This is achieved through the deliberate coloring of both charts and textual information, such as, for instance, the coloring of gender distribution across sectors. Every additional element that is added to any screen adds to the cognitive load of the viewer, thus one needs to carefully review whether each element provides useful information and is necessary (Knaflic, 2015). Therefore, the final rounds consisted of removing unnecessary or distracting objects and fine-tuning the visuals in order to convey the message. Moreover, since excessive information clutters the message, and to maintain a minimalistic design, the tooltip function and hovering text boxes are used to discreetly provide necessary explanations and more detailed information to strengthen contexts of the charts.

Findings and Discussion

We introduce two case studies below to better illustrate how our two target audiences may look to utilize the dashboard alongside the findings for each analysis. But broadly, the dashboard highlights numerous key themes. The first is that overall gender equality across the EU, by our Overall Index measure, has increased in the time period (from 0.75 to 0.77). This comes on the back of progress across 4 of the 7 sub-indices, versus the slight deterioration in the Pay and Economic Sectors indices. The second is that the Care and Decision Makers categories remain the 'low-hanging fruit' for EU Commision members to consider when looking to achieve progress in gender equality, with both sub-indices consistently making up the bottom 2 measures. The third is that longer-term membership of the EU does not necessitate further progress in achieving gender equality - Case Study 2 goes to show that it is generally newer members that make up the top achieving countries by Overall Index. The fourth is that Central Eastern EU states tend to struggle with achieving better outcomes for female decision makers, although this has improved over the time frame. The fifth is that more material progress in gender equality only came about from 2018 onwards - up until then the Overall Index remained consistently at 0.75.

Case Study 1 - Introduction of a Minimum Wage Agreement

As previously discussed, one way in which our target audiences - EU Commision members and EU state policymakers - can use the dashboard, is as an initial insight tool for checking how hypothetical policy interventions could affect progress towards gender equality. An example of this is the introduction of a formal minimum wage agreement in countries that either have no mechanism in place or rely primarily on collective bargaining to set wages (International Labor Organization, 2015). The rationale is simple - “Minimum or living wages are key to reducing the gender pay gap since women are more likely than men to work in the lowest paid jobs” (IndustriALL, 2022) - yet actual outcomes may be less easy to disentangle given the key role of collective bargaining and union powers in particular EU states. If the dashboard is able to give an introductory overview to these sorts of questions in a simple, clear and quick way, it serves a purpose of directing targeted further investigation.

In 2015, Germany introduced its Minimum Wage Act, at the time radically breaking away from a “system of industrial relations and labor market governance” (Eichhorst, 2015, p. 28). It therefore serves as an ideal case study for our target audiences to examine, assessing whether similar policy efforts in other countries can have a positive impact on progress towards gender equality, and specifically towards closing gender pay gaps.

Using the front page of the dashboard as a starting point, hovering over Germany and neighboring states in the years prior to 2015 suggests that, whilst fairly high in absolute terms, its Pay Index relative to others is actually lower. Figures 9 and 10 below show one of these quick comparisons for 2014.

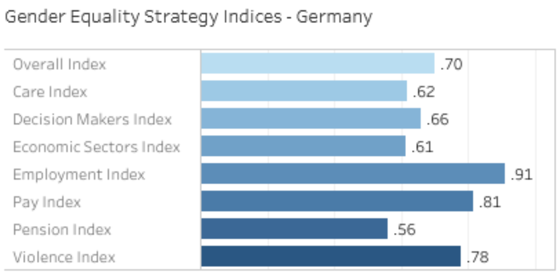


Figure 9: Overall Index / sub-indices for Germany, 2014

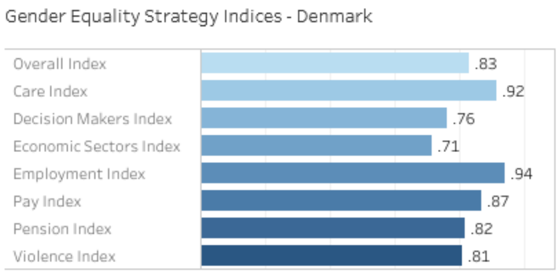
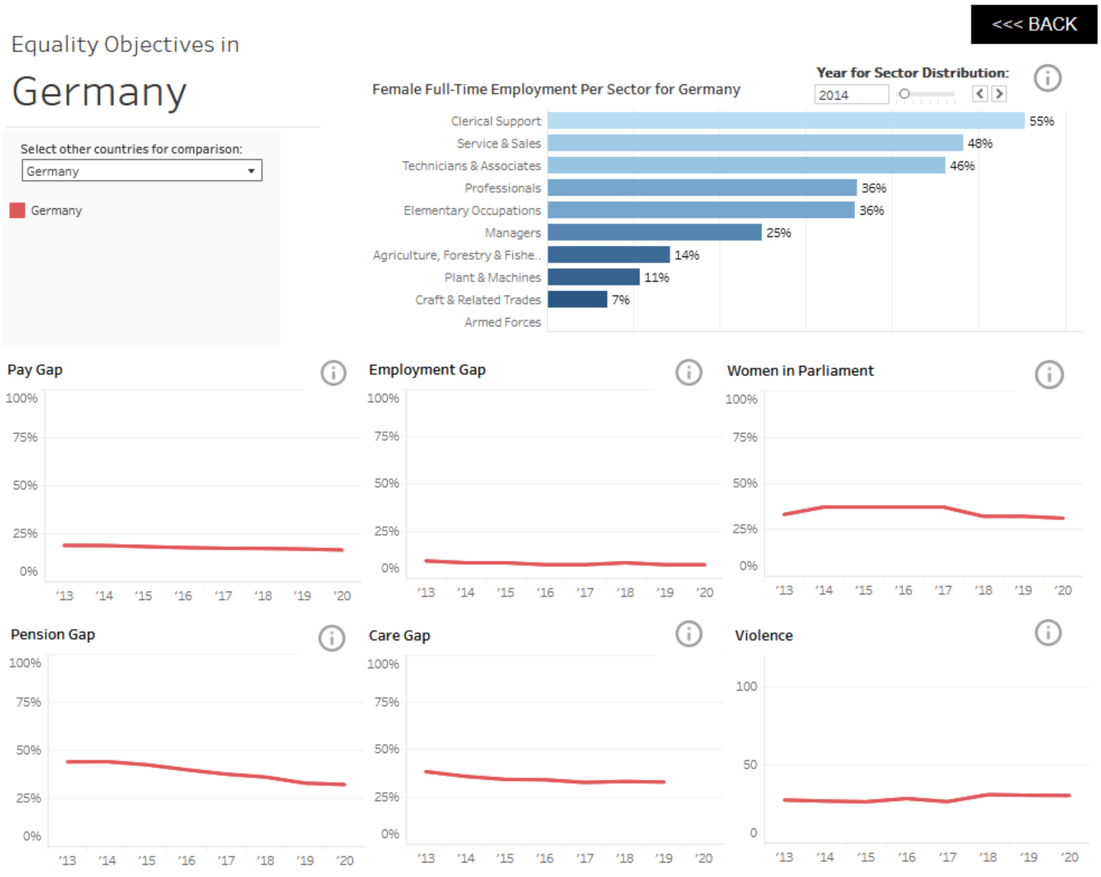


Figure 10: Overall Index / sub-indices for Denmark, 2014

Clicking into Germany on the main map provides the country specific view with greater detail around the underlying numbers driving each sub-index (Figure 11). It can be immediately seen that females made up 55% of Clerical Support roles in 2014 - a clear candidate for a minimum wage agreement. The time series line charts underneath also suggest, at first view, that the pay gap has narrowed by 2%, and the pension gap by 10% since 2015.



Equality Objectives in

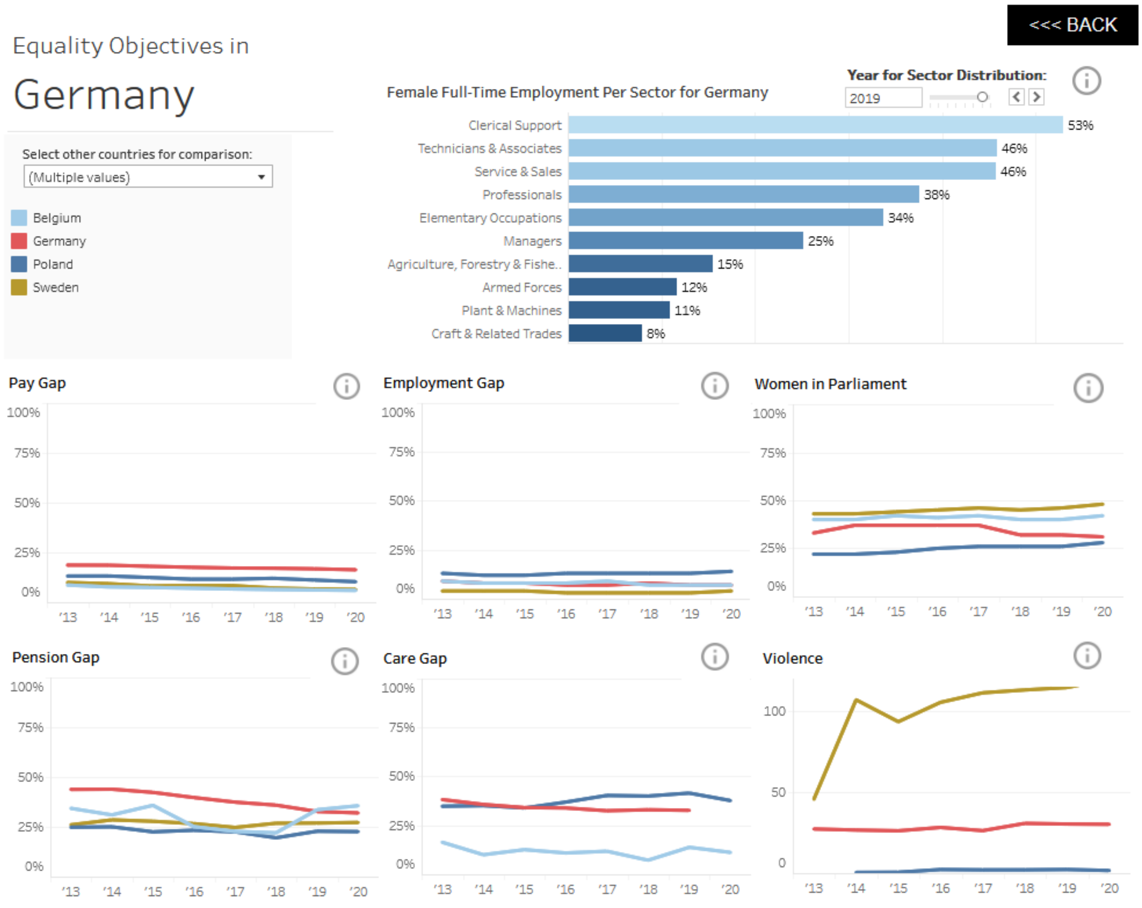
Germany

Select other countries for comparison: Germany

Year for Sector Distribution: 2014

Figure 11: Selective view for Germany, examining the underlying data

Clearly further examination requires comparison versus other countries to check whether these improvements in pay and pension gaps are merely part of a broader shift or indicative of something more specific to Germany. Picking a mixture of neighboring comparison countries with minimum wage agreements (Belgium, Poland) and none (Sweden) allows the user to sift through a number of potential effects (Figure 12). Ultimately the results here suggest that the closing of pay gaps is part of a broader shift in both minimum wage and non-minimum wage countries, as the rate of change over the 2015+ time frame for multiple countries is similar, albeit at different starting levels. The closing of the pension gap in Germany however is much larger, so potential second-order effects that may not initially be obvious from the introduction of a minimum wage agreement can be picked out from this page. For the target audience, this would warrant further investigation with more detailed data sets - but its ultimate purpose of providing quick insights and unearthing more interesting trends is clearly achieved. The primary audience of EU Commision members could use this type of case study to help produce recommendations on the EU level - perhaps the introduction of an EU-wide minimum wage agreement - whilst the secondary audience of EU state policymakers may embed this process into their policy analysis process when weighing alternative implementations (e.g. collective bargaining and union assistance vs minimum wage regulations).



Equality Objectives in

Germany

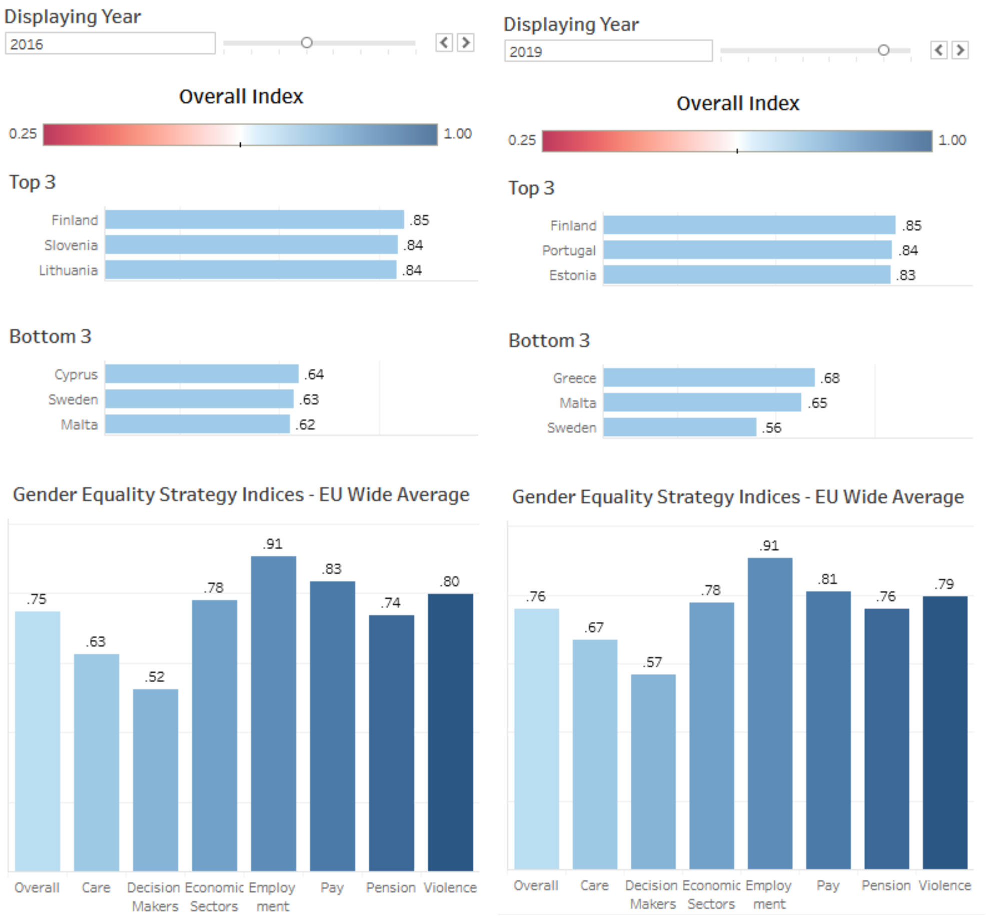
Select other countries for comparison: (Multiple values)

Year for Sector Distribution: 2019

Figure 12: Country comparison - Germany, Belgium, Poland, Sweden

Case Study 2 - Was the European Commission's 2016-2019 Strategic Engagement for Gender Equality Effective? Lessons and Implications

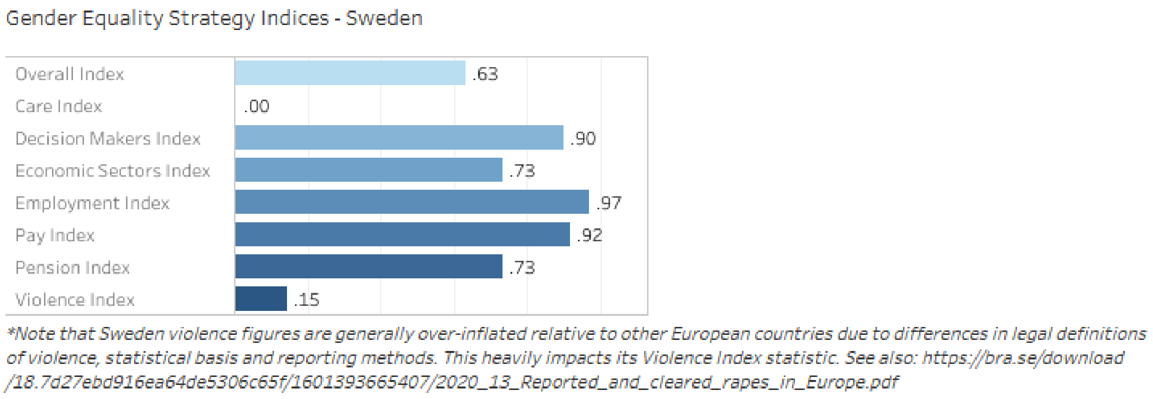
It was earlier discussed that an effective dashboard design for the primary audience, European Commision members, would allow for quick analysis of previous strategies and, by virtue, the formulation of future strategy. Its 2016-2019 Strategic Engagement preceded the 2020-2025 Strategy. Many of its priorities overlap, yet the 2020-2025 iteration expanded upon the objectives by adding the Care category. It is in the same sense that easily observing relationships between sub-indices and countries over time may generate ideas for further priorities going forward. One way our primary target audience may think about strategy formulation is the effectiveness of the 2016-2019 Engagement - which areas it exceeded in, and which areas it made little impact. This begins with an EU-wide analysis on the first page of the dashboard (Figure 13 and 14).



Figures 13/14: EU averages for Overall Index and sub-indices 2016 / 2019

The EU-wide averages of the Overall Index and its sub-indices suggest only minor improvements between 2016 and 2019 - the small increase in the Overall Index from 0.75 to 0.76 was led largely by improvements in the Care and Decision Makers indices, but capped by a worsening in the Pay and Violence indices. Material progress in the Decision Makers and Care categories suggests that the actions taken in the Engagement were successful and should continue to be used. But it can be seen that these are, in a sense, ‘low hanging fruit’ given that they are the two lowest sub-indices overall. Policy actions in the Pay category however possibly need refining. Quickly, policy priorities and successes become clear when framed this way - key for formulating future strategy.

An additional layer of information is provided by the ‘Top 3’ and ‘Bottom 3’ country charts by Overall Index. Finland remains a clear success story, having been ranked 1st in 3 out of the 4 years over this timeframe. Notable also is the prominence of the Baltic states and more recent joiners of the EU - Slovenia and Bulgaria - who frequently feature in the top 3. The other side is the surprising consistency of the bottom 3 performers, of which Sweden and Malta feature across all years. Given the positive progress seen in neighboring countries, the Sweden case stands out. Further investigation when hovering over the country provides the answer (Figure 15).



**Note that Sweden violence figures are generally over-inflated relative to other European countries due to differences in legal definitions of violence, statistical basis and reporting methods. This heavily impacts its Violence Index statistic. See also: https://bra.se/download /18.7d27ebd916ea64de5306c65f/1601393665407/2020_13_Reported_and_cleared_rapes_in_Europe.pdf*

Figure 15: Hover information over Sweden, Aggregate View. Footnote below clarifying information.

The informational footnote explains the poor violence figures resulting from Sweden's broader definitions of crime and reporting methods. Comparing other sub-indices to the EU wide average displayed in the bottom-right corner, it is clear that Sweden is generally outperforming on the Decision Makers, Employment and Pay fields, so EU Commission members may therefore wish to weigh the Violence index accordingly in their assessment.

Informed by what is generally driving the top and bottom performers, Commission members may wish to drill down into the underlying data in the country view and compare against neighbors or regional counterparts. When developing future strategy, country or regional specific actions may be possible - for example, a clear under-representation of women in Parliament across parts of Eastern Europe becomes apparent under further investigation (Figure 16).

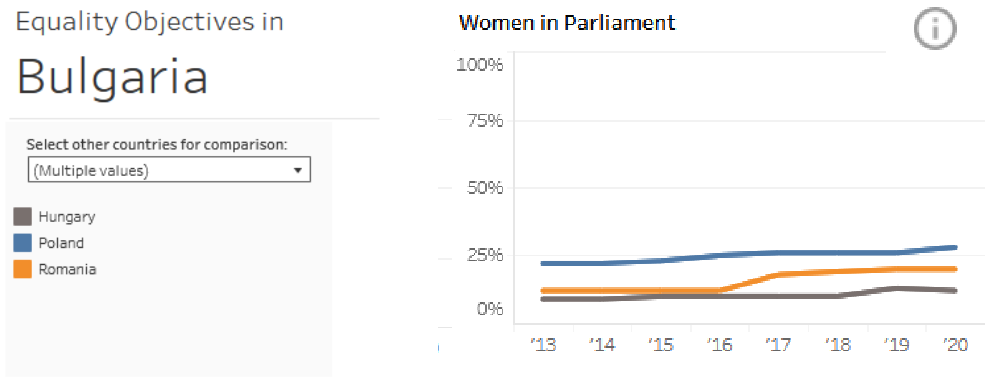


Figure 16: Eastern Europe comparison via Selective View - Women in Parliament figures low across the board

This case study suggests that lessons from EU strategies can be learnt and help to inform future actions. Clearly more detailed analysis is required, but as a starting point for policy formulation, the dashboard has clear benefits to our target audience.

Conclusion

The discussion in this report has worked through how the EU Gender Equality Dashboard can be effectively utilized by two key audiences - European Commission members formulating broader strategy around gender equality, and EU state policymakers enacting this strategy within their own countries. It has implications for where and how to target action on the EU group and country level by allowing for quick comparisons of progress both over time and between states. Furthermore, the generation of a unique index and sub-indices adds another layer of insight by summarizing complex data into something easily interpreted by target decision makers. Since the data is bespoke to the Gender Equality Strategy 2020 - 2025, it also offers something new and tailored to policy efforts relative to existing measures of progress. The dashboard allows for analysis of previous strategy efforts, hypothetical policy interventions and the ready extraction of themes and trends in gender equality.

Given the many data sets used in creating this dashboard, this report explained the systematic approach taken to cleaning and standardizing materially different sources. Whilst the pre-processing structure remained broadly similar, individual decisions were made around handling null values, generating statistics from raw data and the creation of an Overall Index. Discussion around the design of the dashboard explained how, from the outset, creation of an ‘Aggregate’ and ‘Selective’ view, producing a tool useful to both of the target audiences was adhered to. However, establishing the visual identity and logic connecting these two views was the outcome of an iterative process.

Issues and Potential Further Work

In attempting to summarize a complex topic in 7 primary data points, issues and limitations clearly arise. Capturing the essence of multi-faceted objectives like ending gender-based violence is not the same as measuring the reduction in pay gaps, for example. Even targets that appear more clear-cut are still layered - age, ethnicity, income level amongst a multitude of other factors ideally need to be taken into account when creating a gender equality strategy. Data collection in certain areas is still only just beginning to materialize, meaning insufficient sources are available for some categories such as violence and care. Even the Commission itself admits that “The EU needs comprehensive, updated and comparable data for policies on combating gender-based violence to be effective." (European Commission, 2020). This has implications for the data sets that do exist - null values were frequent in these areas and measures, such as the averaging of multiple violence statistics, were taken to handle the breadth of the target and missing data. Infrequent - mostly annual - measurements also forced the granularity in terms of time frame to be limited to yearly figures across our index calculation. Null values in sub-indices, whilst also handled appropriately do limit some of the informational value that the Overall Index can provide when comparing countries.

By its very nature, data taken from across 27 countries draws up comparability risks. This is especially prominent where different authorities are involved in its collection and processing. That effect is clearest in the Sweden case, where a broad interpretation of the law artificially increases its violence incidence statistics. While this is flagged to the user, it raises questions about the data set more broadly.

Further development of the dashboard would benefit from greater granularity in areas where this is possible. Regional analysis within a number of countries was a possibility and something the second target audience - EU state policymakers - could benefit from when implementing actions. This would allow for a more targeted approach that could perhaps make the overall process more effective. A further view examining country-level performance versus threshold levels of progress across sub-indices could also enrich the process of future strategy formulation by the primary target audience, EU Commission members. As it stands, the dashboard provides a lookback on previous years, but ideally as data is released it would incorporate the latest figures (i.e. 2021 and 2022 for all data sets) to establish the impact of the current Gender Equality Strategy. However, as earlier discussed, progress in achieving gender equality is a long-term goal, taking time to materialize; the 0.02 increase in the EU-wide Overall Index over 2013 to 2020 reflects exactly that point.

References

Eichhorst, W. (2015). Understanding the German minimum wage. *Samfundsøkonomen, 2015(1),* 28-31.

European Commission. (2020, March 5). *A Union of Equality: Gender Equality Strategy 2020-2025*. European Union. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52020DC0152

European Commision. (2020b). *Gender Equality Strategy. Achievements and key areas for action.* https://commission.europa.eu/strategy-and-policy/policies/justice-and-fundamental-rights/gender-equality/gender-equality-strategy_en

European Institute for Gender Equality. (2017). *Home*. Gender Statistics Database. https://eige.europa.eu/gender-statistics/dgs/indicator/bpfa_f_offic_f12__bpfa_f12a/metadata

European Institute for Gender Equality. (2022). *Gender Equality Index.* EIGE Europa. https://eige.europa.eu/gender-equality-index/2022/EU

European Institute for Gender Equality. (2022b). *Indicator: National parliaments: presidents and members | Gender Statistics Database | European Institute for Gender Equality*. European Institute for Gender Equality. https://eige.europa.eu/gender-statistics/dgs/indicator/wmidm_pol_parl__wmid_natparl/metadata

European Institute for Gender Equality. (2022c). *Gender Equality Index - Violence.* EIGE Europa. https://eige.europa.eu/gender-equality-index/2022/domain/violence

Eurostat. (2021). *Employment and unemployment (Labour force survey) (employ)*. European Commission. https://ec.europa.eu/eurostat/cache/metadata/EN/employ_esms.htm#coher_compar1667912644217

Eurostat. (2022). *Intentional homicide and sexual offences by legal status and sex of the person involved - number and rate for the relevant sex group*. EU. https://ec.europa.eu/eurostat/databrowser/view/crim_hom_soff/default/table?lang=en

Eurostat. (2022b). *Employment and activity by sex and age - annual data*. EU. https://ec.europa.eu/eurostat/databrowser/view/LFSI_EMP_A__custom_3997920/default/table?lang=en

Eurostat. (2022c). *Full-time and part-time employment by sex, age and occupation (1 000)*. EU. https://ec.europa.eu/eurostat/databrowser/view/LFSQ_EPGAIS__custom_4004302/default/table?lang=en

Eurostat. (2022d). *Gender pay gap in unadjusted form by NACE Rev. 2 activity - structure of earnings survey methodology*. EU. https://ec.europa.eu/eurostat/databrowser/view/EARN_GR_GPGR2__custom_4010217/default/table?lang=en

Eurostat. (2022e). *Gender pension gap by age group – EU-SILC survey*. EU. https://ec.europa.eu/eurostat/databrowser/view/ILC_PNP13__custom_4016755/default/table?lang=en

Eurostat. (2022f). *Inactive population due to caring responsibilities by sex*. EU. https://ec.europa.eu/eurostat/databrowser/view/sdg_05_40/default/table?lang=en

Few, S. (2006). *Information Dashboard Design: The Effective Visual Communication of Data*. O'Reilly.

Human Development Report. (2016). *Technical Notes.* *https://hdr.undp.org/sites/default/files/data/2020/hdr2016_technical_notes.pdf*

IndustriALL. (2022). *Achieving pay equity through collective bargaining.* *https://www.industriall-union.org/feature-achieving-pay-equity-through-collective-bargaining*

International Labour Organization. (2015). *Minimum wages through collective bargaining.* ILO. https://www.ilo.org/global/topics/wages/minimum-wages/setting-machinery/WCMS_436112/lang--en/index.htm

Knaflic, C. (2015). *Storytelling with Data: A data visualization guide for business professionals*. John Wiley & Sons.

Mishra, S., & Nathan, H.S.K. (2013). *Measuring Human Development Index: The old, the new and the elegant.* (IGIDR Working paper WP-2013-020). https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=3694b5471c763096b4a73d6231d9651c89423f93

Organisation for Economic Co-operation and Development. (2022). *Employment : Gender wage gap*. OECD Statistics. https://stats.oecd.org/index.aspx?queryid=54751

The World Bank Group. (2023). *All Indicators*. World Bank Gender Data Portal. https://genderdata.worldbank.org/indicators/

The World Bank Group. (2023b). *Home*. Gender Data Availability. https://genderdata.worldbank.org/data-availability/?tab=availability&year-range=10

Wickham, H. (2014). Tidy Data. Journal of Statistical Software, 59(10), 1–23. https://doi.org/10.18637/jss.v059.i10

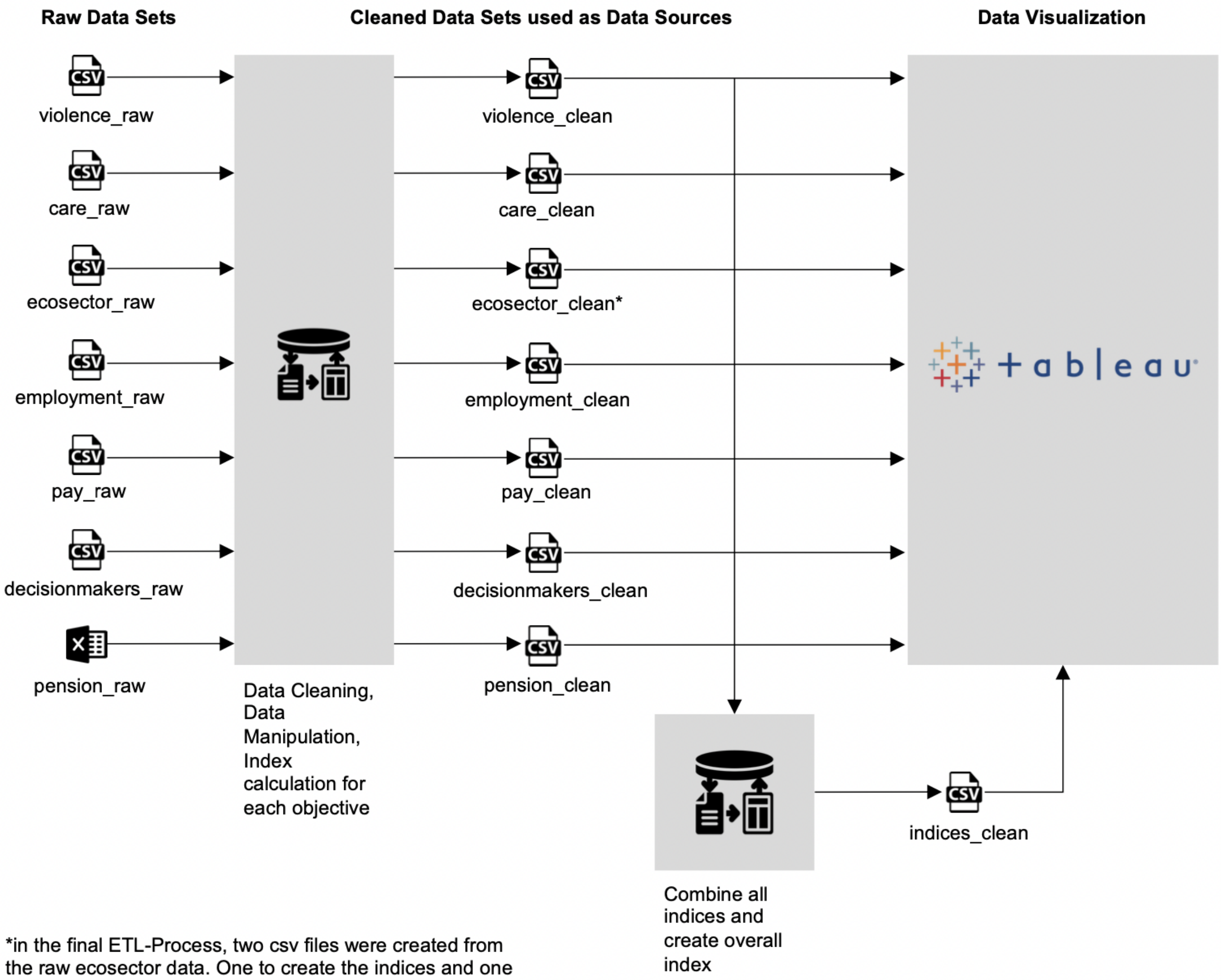
World Health Organization. (2022). *“Human Development Index”.* *https://www.who.int/data/nutrition/nlis/info/human-development-index#:~:text=It%20was%20created%20to%20re,a%20country%2C%20not%20economic%20growth.*

Appendix

Appendix 1: Identified datasets, represented key objectives and comments

Dataset incl. Data Source	Represented Key Objective	Comment
Intentional homicide and sexual offences by legal status and sex of the person involved - number and rate for the relevant sex group (Eurostat, 2022)	Ending gender-based violence	Mixed accuracy by different criminal administrations. Comparability mixed both geographically and over time due to differences in legal systems, recordings.
Employment and activity by sex and age - annual data (Eurostat, 2022b)	Closing gender gaps in the labor market	High reliability
Full-time and part-time employment by sex, age and occupation (1000) (Eurostat, 2022c)	Achieving equal participation across different sectors of the economy	EU's Labour force survey (LFS) provides overall high quality statistics from national LFS.
Gender pay gap in unadjusted form by NACE Rev. 2 activity - structure of earnings survey methodology (Eurostat, 2022d)	Addressing the gender pay gap	Somewhat mixed geographical comparability, but good over time
Gender pension gap by age group – EU-SILC survey (Eurostat, 2022e)	Addressing pension gaps	High reliability and only one missing data value, good geographical and time comparability
Inactive population due to caring responsibilities by sex (Eurostat, 2022f)	Closing the gender care gap	Comparability over time of <3 data points, but good across geographies due to single survey measure
National parliaments: presidents and members (European Institute for Gender Equality, 2022)	Achieving gender balance in decision-making and in politics	

Appendix 2: Outline of the ETL-Process



*in the final ETL-Process, two csv files were created from the raw ecosector data. One to create the indices and one to display the data for each sector in Tableau

Appendix 3: Example of the applied process guideline to one of the datasets

```
# Import packages
import pandas as pd
import numpy as np

# Reading in raw csv files
raw_df = pd.read_csv('./datasets_raw/Employment and activity by sex and age - annual
data_eurostat_2022.csv')
print(raw_df.head())
print(raw_df.columns)

# Melting data in the correct format
   # No need to melt the data in this dataset

# 1. Data Cleaning Step 1
# 1.1 Identify and delete columns that are not necessary for further process
raw_df.drop(columns=['DATAFLOW', 'LAST UPDATE'], axis=1, inplace=True)
print(raw_df.head())

# 1.2 Explore each column and there unique values, especially if categorical data AND
Decide how to proceed with this data
# Only one value in column 'freq' (frequency) --> annually, so that we can drop this
column as well
print(raw_df['freq'].unique())
raw_df.drop(columns=['freq'], axis=1, inplace=True)
print(raw_df.head())

# Indices / indic_em
# ACT: Persons in the labour force
# EMP_LFS: Total employment (resident population concept)
# We want to see data of the total employment and not only for active persons in the
labour market
print(raw_df['indic_em'].unique())

# Sex
# F: Female
# M: Male
# T: Total
print(sorted(raw_df['sex'].unique()))

# Age
# Data has different age groups
print(raw_df['age'].unique())

# Unit
# PC_POP: Percentage of total population
# THS_PER: Thousand persons
# We are going to work with the percentage of total population
print(raw_df['unit'].unique())

# Geo
# Working with the countries that are inside the EU27 csv from addtional data
print((raw_df['geo'].unique()))

# TIME_PERIOD
# Time is from 2003-2021, but maybe probably not in all countries
print(sorted(raw_df['TIME_PERIOD'].unique()))

# OBS_VALUE
# mixed values percentage and total values --> getting rid of the total values through
the unit
print(sorted(raw_df['OBS_VALUE'].unique()))

# OBS_FLAG
# Deciding in a later stage what to do with the flagged rows
# b: break in time series
# d: definition differs
# nan: no flag
print((raw_df['OBS_FLAG'].unique()))

# Clean dataset STEP 1
# indic_em = EMP_LFS
# unit = PC_POP
# Drop columns after filtering
clean_df = raw_df.loc[(raw_df['indic_em'] == 'EMP_LFS')].copy()
clean_df.drop(columns=['indic_em'], axis=1, inplace=True)
print(clean_df.head())

# 1.3 Rename country column with help of the EU27.csv in additional data
# Create dictionary of EU27 countries to apply map function
countries_df = pd.read_csv("./additional_data/EU27_COUNTRY_LIST.csv")
countries = dict(zip(countries_df['Initial'].str.strip(), countries_df['Country']))
print(countries)

# Apply map function to rename the countries
```

```
new_country_column = clean_df['geo'].map(countries)
clean_df.loc[:,'geo'] = new_country_column
print((clean_df.head()))

# renaming columns
clean_df.rename(columns={"sex": "Sex", "age" : "Age", "geo": "Country", "TIME_PERIOD":
"Year", "OBS_VALUE": "Total Employment in %", "OBS_FLAG" : "Flag"}, inplace=True)
print((clean_df.head()))

# reorder columns
clean_df = clean_df[['Country', 'Year', 'Sex', 'Age', 'Flag', 'Total Employment in
%','unit']]
print((clean_df.head()))

# Flags
# Information to Flags, various reasons:
# https://ec.europa.eu/eurostat/cache/metadata/en/lfsi_esms.htm
# "Overall, comparability over time is considered as high."
# 'b's are most likely a change in statistical method:
# "Methodological improvements in the underlying sampling design or changes in
nomenclatura can lead to breaks in the time series."
# 'd' only in year 2021 for Spain and France
# Therefore, we can get rid of the Flag column
print(clean_df['Flag'].unique())
print(clean_df.loc[clean_df['Flag'] == 'd'])
print(clean_df.loc[clean_df['Flag'] == 'b'])
clean_df.drop(columns=['Flag'], axis=1, inplace=True)


# 2. Data Cleaning Step 2
# 2.1 Explore data with df.info() /df.describe() and clean df if necessary
# Null Values in Country column are those countries that do not belong to the EU27
# Therefore removing all rows with Country "NaN"
# Now, there are no null values in the df anymore
print(clean_df.info())
clean_df2 = clean_df[clean_df['Country'].notna()].copy()
print(clean_df2.info())

# 2.2 Compare if each country has the same number of rows
# Each item of the countries list has 234 rows
countries = list(countries_df['Country'])
for c in countries:
       print('Country: ' + str(c) + ': '+ str((clean_df2['Country'] == c).sum()))
```

```
# 2.3 Compare if each year has the same number of rows
# The years 2003 - 2008 only have 18 rows
# The years 2009 - 2021 have all 504 rows
# We want to work only with years, where each country has the same data to compare them
completely
# Therefore, we are removing all rows with the years 2003-2008
for i in range(1990,2025):
   print('Year: ' + str(i) + ': '+ str((clean_df2['Year'] == i).sum()))
clean_df2.drop(clean_df2[(clean_df2['Year'] < 2009)].index, inplace=True)
print(clean_df2.info())
print(clean_df2.describe())
print(clean_df2.head())

# 3. Further individual cleaning
# Calculate percentage values
# Create one df with percentage values
# Create another df with total values --> clean_df4
clean_df3 = clean_df2.loc[(clean_df2['unit'] == 'PC_POP')].copy()
clean_df4 = clean_df2.loc[(clean_df2['unit'] == 'THS_PER')].copy()
clean_df3.drop(columns=['unit'], axis=1, inplace=True)
clean_df4.drop(columns=['unit'], axis=1, inplace=True)

clean_df3['Total Employment in %'].astype('float')
clean_df3['Total Employment in %'] = clean_df2['Total Employment in %'].div(100).round(2)
#print(clean_df3.head())

# Setting index to perform mathematical operations
clean_df3.set_index(['Country', 'Year','Age'], inplace=True)
clean_df4.set_index(['Country', 'Year','Age'], inplace=True)

# Split datasets into the Genders
fem_df = clean_df3[clean_df3['Sex'] == 'F'].copy()
mal_df = clean_df3[clean_df3['Sex'] == 'M'].copy()
tot_df = clean_df4[clean_df4['Sex'] == 'T'].copy()

# Column is not needed anymore
fem_df.drop(columns=['Sex'], axis=1, inplace=True)
mal_df.drop(columns=['Sex'], axis=1, inplace=True)
tot_df.drop(columns=['Sex'], axis=1, inplace=True)

# Calculate the employment gap between man and woman
fem_df['Employment Gap in %'] = mal_df['Total Employment in %'].sub(fem_df['Total
Employment in %']).round(2)
```

```
# Add the total number of employed persons for each row
fem_df['Employed Persons in Thousands'] = fem_df.index.map(tot_df['Total Employment in
%'])
fem_df['Employed Persons in Thousands'] = fem_df['Employed Persons in
Thousands'].astype(int)

# Drop unnecessary columns
#print(fem_df[fem_df['Employment Gap in %'] == 0.0])
fem_df.drop(columns=['Total Employment in %'], axis=1, inplace=True)

# Reset index after finishing mathematical operations
fem_df.reset_index(inplace=True)
print(fem_df.head())

# Calculating index: (Actual value - worst value)/(Best value - worst value)
# Taking the max. employment gap from age group 15-64 since this is the age group where
we are taking the index to calculate the overall index
best_value = 0
worst_value = 1
fem_df['IndexValueEmployment'] = (fem_df['Employment Gap in
%']-worst_value)/(best_value-worst_value)
#print(fem_df[fem_df['Age'] == 'Y15-64']['IndexValueEmployment'].max())

# Drop European Union Stats
i = fem_df[(fem_df['Country'] == 'European Union')]
fem_df.drop(i.index, inplace=True)

# 5. Save cleaned dataframe in folder datasets_cleaned
fem_df.to_csv('./datasets_cleaned/Employment by sex and age.csv')
```